

IMAGE ARTIFACT REDUCTION USING A NEURAL NETWORK

BACKGROUND

[0001] Data compression is often used for reducing the cost of storing large data files on computers as well as reducing the time to transmit large data files between computers. In the so-called "transform methods" data is transformed into coefficients that represent the data in a frequency domain. Coefficients may be quantized (lossy compression), and redundancy in the quantized coefficients may then be reduced or eliminated (lossless compression).

[0002] JPEG is a standardized image compression algorithm. JPEG compression of an image includes dividing the image into a grid of non-overlapping 8x8 blocks of pixels, and independently coding each pixel block. The coding of each pixel block includes taking a two-dimensional Discrete Cosine Transform (DCT) to obtain an 8x8 block of DCT coefficients; and quantizing the DCT coefficients. The quantization exploits the following: the low frequency DCT coefficients contain most of the image energy; sensitivity limits of the human visual system vary with spatial frequency (e.g., small high frequency changes are perceived less accurately than small low frequency changes); and the human visual system is much more sensitive to high frequency variations in luminance than similar variations in color.

[0003] The image may be reconstructed by performing an inverse DCT transform on the quantized coefficients. Because the coefficients are quantized, the reconstructed image does not contain all of the information of the original image (that is, the image prior to compression). Consequently, the reconstructed image is not identical to the original image.

[0004] Moreover, the reconstructed image can contain artifacts that were not present in the original image. For example, compare Figures 2a and 2b, which are images of a sliced fruit against a textured background. Figure 2a shows the sliced fruit and textured background prior to JPEG compression. Notice the gradual change in color of the textured background, and the crisp edges between the sliced fruit and the background.

[0005] Figure 2b shows the sliced fruit and textured background after the image was JPEG-compressed and thereafter decompressed. Now notice the background texture of the decompressed image. Instead of texture, there appear groups of blocks of different shades (each 8x8 DCT block is smoothed to a single shade). These artifacts are referred to as “blocking” artifacts. In addition, the edges of the sliced fruit are no longer crisp. Echoes or shadows appear at the edges. These artifacts at the edges are referred to as “ringing” artifacts.

[0006] The blocking and ringing artifacts can degrade image quality. They are especially prominent if the JPEG-compression was performed at a low bit rate (i.e., a highly compressed image).

[0007] It is desirable to reduce the artifacts in decompressed images.

SUMMARY

[0008] According to one aspect of the present invention, a neural network is trained and used to reduce artifacts in spatial domain representations of images that were compressed by a transform method and then decompressed. For example, the neural network can be trained and used to reduce artifacts such as blocking and ringing artifacts in JPEG images.

[0009] Other aspects and advantages of the present invention will become apparent from the following detailed description, taken in conjunction with the accompanying drawings, illustrating by way of example the principles of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Figure 1 is an illustration of a method of performing artifact reduction in accordance with an embodiment of the present invention.

[0011] Figures 2a-2c are images of the same scene.

[0012] Figure 3 is an illustration of a method of training a neural network to perform artifact reduction in accordance with an embodiment of the present invention.

[0013] Figure 4 is an illustration of an exemplary neural network architecture.

[0014] Figure 5 is an illustration of a method of training a neural network in accordance with another embodiment of the present invention.

[0015] Figure 6 is an illustration of an apparatus in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

[0016] As shown in the drawings for purposes of illustration, the present invention is embodied in methods of training and using a neural network to reduce artifacts in a JPEG image. The JPEG image refers to a digital image (the "original" image) that was compressed in accordance with the JPEG standard and then decompressed. Thus the JPEG image is represented in the spatial domain. Although the present invention is described in connection with JPEG, it is not so limited. The present invention may be used in connection other transform methods that utilize lossy compression.

[0017] In the following paragraphs, the training and use of the neural network will be described in connection with a grayscale JPEG image. Later on, training and use of full color JPEG images will be addressed.

[0018] Reference is made to Figure 1. A grayscale JPEG image 110 is inputted to a neural network 112, which has already been trained to perform artifact reduction. For each pixel being processed, an input vector 111 is formed. The input vector 111 may include the pixel being processed (indicated by the "X"), and a plurality of pixels (indicated by circles) in a neighborhood of the pixel being processed. For example, each input vector 111 could be formed from a 3x3 window containing the pixel being processed and eight neighboring pixels (as shown in Figure 1), or each input vector 111 could be formed from a 5x5 window containing the pixel being processed and twenty four neighboring pixels. Prior to processing, the input vector 111 may be scaled by subtracting the value of the pixel being processed from the value of each pixel in the input vector 111. If this is done, the scaled input vector 111 need not include the pixel being processed, which would have a value of zero.

[0019] The neural network 112 processes the input vectors 111 independently of one another. Each input vector 111 is processed to produce a single pixel 113 in an output image 114. This output pixel 113 has the same spatial position as the pixel being processed. Thus each pixel value in the output image 114 is predicted from a contextual window of the JPEG image 110. If artifacts such as blocking and ringing appear in the JPEG image 110, those artifacts are generally reduced in the output image 114.

[0020] The input vectors 111 may be processed serially by a single processor system. However, since the input vectors are processed independently of one another, a multi-processor system can process a plurality of input vectors 111 in parallel.

[0021] Reference is now made to Figures 2a-2c. As discussed above, Figure 2a shows the original image of a scene including sliced fruit against a textured background, and Figure 2b shows the same scene after the image has been JPEG-compressed and thereafter decompressed. The decompressed image contains blocking artifacts in the background and ringing artifacts at edges of the sliced fruit.

[0022] Figure 2c shows the output image produced by the trained neural network 112 (this is an actual image produced by a neural network that was reduced to practice). The neural network 112 does not eliminate the artifacts, but it does reduce the artifacts. The output image still contains blocking artifacts, but those artifacts are smoothed. Notice the background, which still contains blocks of different shades. The output image still contains ringing artifacts, but the ringing artifacts are not as noticeable.

[0023] Training of the neural network 112 will now be described. The neural network 112 is defined by its nodes, connections, and connection weights. A weight vector is the vector of connection weights between each pair of connected nodes in the neural network 112. Training involves optimizing these weights so as to reduce the error between the output image 114 and the original image 110.

[0024] Reference is now made to Figure 3, which illustrates a method of training the neural network 112. A JPEG image is generated by compressing an original image according to the JPEG standard and then decompressing the

compressed image (310). The JPEG image and the original image form a training pair.

[0025] The JPEG image is inputted to the neural network (312). An input vector is formed for each pixel, and each input vector is inputted to the neural network 112.

[0026] The neural network forward-propagates each of the input vectors to compute values of the nodes (314). The connection weights are used to compute these node values. During forward propagation, values from the hidden and output nodes may be obtained by computing the network-weighted sum in a progressive manner. For example, the input to a first hidden node is the weighted sum of the inputs from the given input pattern. The weights used for the weighted sum are the current values of the connections between the inputs and the first hidden unit. The output of the first hidden node is a non-linear activation function (e.g., a hyperbolic tangent) of the input. Once this is computed, the input to the second hidden unit is computed as the appropriate weighted sum of the inputs and the output of the first hidden node, and so forth. The output values from the output nodes represent the current network prediction for the output image (316).

[0027] Errors between pixels of the output image and corresponding pixels of the original image are determined (318). For example, an error image may be formed by subtracting the output image from the original image.

[0028] Derivatives of the errors are computed with respect to the output image (320). The derivatives may be computed in a conventional manner, for example by using sum-squared errors (SSE). In the alternative, the derivatives may be computed from spatial errors as disclosed in assignee's U. S. Serial No. 10/600,671 filed June 20, 2003, which is incorporated herein by reference. Using spatial errors, the derivative for a predicted pixel in the output image is a function of differences between predicted values in a spatial neighborhood of the output image and the corresponding values in the original image.

[0029] If a spatial error measure is used, the spatial error measure may also involve reducing the clustering of undesirable spatial patterns of errors (328). This step is described in U.S. Serial No. 10/600,671.

[0030] Once the derivatives for the pixels in the output image have been generated, back-propagation is performed to compute error gradients (322). The error gradients may be computed as derivatives of the network output with respect to the network weights. The back-propagation may be performed in a conventional manner. For example, using the chain rule of differentiation, the derivative of the network error may be expressed with respect to the network weights as a product of the derivatives of the network error with respect to the network output and the derivatives of the network output with respect to the network weights.

[0031] The error gradients are used to adjust the node weights to reduce the network errors (324). This may be done in a conventional manner. The error gradients may be used iteratively to find weights that result in a lower error. For example, a gradient descent optimization algorithm such as BFGS or a simpler gradient descent algorithm may be used.

[0032] Second and subsequent iterations may then be performed until a stopping criteria is reached (326). For each iteration (314-324), an output image is generated from the JPEG image and the adjusted weights, errors are computed, derivatives are computed from the errors, back-propagation is performed, and node weights are further adjusted.

[0033] The stopping criteria may be one of the following, or a combination of the following (the following stopping criteria is exemplary, not exhaustive):

- (1) The neural network error derivative is within a threshold.
- (2) The neural network error is within a threshold.
- (3) The neural network has completed a maximum number of training iterations.
- (4) The neural network has achieved a spatial error minimum when evaluated on an independent validation set.

[0034] The neural network is not limited to any particular architecture. A general feedforward architecture, a single-layer or multi-layer network, or another architecture may be used.

[0035] An exemplary feed-forward neural network architecture may have linear output nodes and tanh activation in the hidden nodes. Each input node is

connected to each hidden and output node, and the connection weight between the i^{th} and j^{th} nodes is represented as $w_{i,j}$. The hidden nodes are ordered, and each hidden node is connected to each subsequent hidden node and to each output node. The first input is a bias input with a constant value of one. Each network node has an index, with index 0 being the bias input, indices 1 to N_{in} corresponding to the input nodes, indices $N_{\text{in}}+1$ to $N_{\text{in}}+N_{\text{hid}}$ corresponding to the hidden nodes, and indices $N_{\text{in}}+N_{\text{hid}}+1$ to $N_{\text{in}}+N_{\text{out}}$ corresponding to the output nodes.

[0036] The output function for the i^{th} node (input, output or hidden) may be represented in terms of previous hidden and input nodes as follows:

$$a_i = \tanh\left(\sum_{j=0}^{i-1} w_{i,j} x_j\right)$$

where a_i represents the i^{th} output node. Since the output nodes are linear, the output of the i^{th} node may be represented in terms of previous output, hidden and input nodes as follows.

$$a_i = \sum_{j=0}^{i-1} w_{i,j} x_j$$

[0037] A generalized feedforward network of any given size can mimic any layered architecture with an equivalent number of total hidden nodes. For example, the feedforward neural network 410 illustrated in Figure 4 has three inputs 412, three hidden nodes 414, and a single output node 416. In practice very few hidden nodes 414 are needed to provide satisfactory results. In particular, neural networks with as few as ten hidden nodes 414 can yield satisfactory results.

[0038] Input and output values of the neural network may be coded to improve the neural network accuracy. The coding may be performed to fit the input and output values within a range (e.g., $[-1,1]$). This is done to better suit the dynamic range of the activation functions, and also to minimize the dimensionality of the input/target space. Exemplary coding schemes include simple coding, relative coding, and scaled coding.

[0039] In simple coding, each input value is scaled to the range $[0,...,1]$ by dividing by 255 (for eight-bit values). This transformation ensures that the network

inputs and outputs always fall between 0 and 1. Each output value is converted back to a value within the range $[0, \dots, 255]$ by multiplying by 255.

[0040] In relative coding, simple coding is performed on each input vector, and the value of the pixel being processed is subtracted from all values in the input vector. The scaled input vector need not include the pixel being processed, since its value was reduced to zero. Inputs and outputs are in the range $[-1, \dots, 1]$. Each output value is then converted back to a value within the range $[0, \dots, 255]$ by $\text{CLIP}(\text{network_output} \cdot 255 + \text{dc})$, where dc is the input value of the pixel being processed (i.e., the central pixel in the input window), and the operation $\text{CLIP}(x)$ returns a 0 if $x < 0$, 255 if $x > 255$, and x otherwise.

[0041] The relative coding makes it easier for the neural network to recognize edges and features. This has the effect of adjusting the DC level and making edges the same, regardless of their grey level.

[0042] In scaled coding the relative coding is computed. Then the inputs and outputs are scaled by a dynamic range of the inputs, so the inputs are usually "stretched" to $[-1, \dots, 1]$. The output is not bounded. Thus the dynamic range of the output pixels may be larger than the dynamic range of the input pixels. Unlike relative coding, scaled coding produces edges that have similar dynamic ranges. As a benefit, the neural network only learns about the shape of edges, and not edge height. For example, in relative coding, an edge with a difference of 70 gray levels would look significantly different than the same edge with a difference of 150 gray levels. Using scaled coding, the neural network can recognize the edge without concern about the height. The scaling factor can be clipped to help prevent noise appear as structure to the neural network, while transforming edges so they appear similar to the neural network.

[0043] The following table demonstrates the range differences between these coding types.

Type	Inputs	Outputs
Simple	$[0, \dots, 1]$	$[0, \dots, 1]$
Relative	$[-1, \dots, 1]$	$[-1, \dots, 1]$
Scaled	$[-1, \dots, 1]$	unlimited

[0044] The training described above is performed with a single JPEG image and its corresponding original image. However, the training is not so limited to a single pair of images. The neural network can be trained on multiple (e.g., twenty five) pairs of images. The image pairs in the training set may include a reasonable collection of good and hopefully representative images with various "types" of images, such as people, landscapes, man-made objects, still life, text, etc. The JPEG images might have been compressed at different quality factors (e.g., one each for quality factors of 20, 40, 60, 80 and 90).

[0045] The neural network may be fully trained on one image in the pair, before being supplied with the next image in the set. In the alternative, the training may be performed over multiple runs. For example, a first training run involves only a small set of training images and results in the neural network being near the desired point. A second training run based on a larger number of input images is then performed until the neural network satisfies a stopping criteria.

[0046] More than one pixel could be processed at a time. For example, the neural network could process pixels in 3x3 blocks when given a 7x7 input block.

[0047] The neural network could be trained using an optimization function (e.g., using a non-gradient based training algorithm such as simulated annealing or genetic algorithms) that does not require or utilize error gradients or derivatives. Such training would be based on the error function.

[0048] Reference is made to Figure 5. Non-gradient based training may include the following steps. Pixel errors between a JPEG image and an original image are determined (510). Undesirable spatial patterns of errors may be penalized (512), as described in U.S. Serial No. 10/600,671. Errors are reduced by generating a new weight vector using the non-gradient based training algorithm such as a non-gradient descent algorithm (514).

[0049] The neural network can be trained to recognize inputs in addition to the JPEG image. For example, the input vectors may include x- and y-offsets. These offsets indicate the distance (in pixels, for example) of the pixel being processed from a block boundary (that is the boundary of a DCT block). Many JPEG artifacts are

more apparent and common near block boundaries. For example, in relatively flat regions, the blocking artifacts are caused by abrupt transitions in the DC value at block boundaries. If the network knows that a pixel is at a block boundary, then it is more likely to guess that such an edge/transition is likely to be an artifact that should be smoothed, rather than a true edge, which should be preserved

[0050] The input vectors are not limited to 5x5 neighborhoods. A larger neighborhood results in a greater dependency on neighbors. Larger neighborhoods take longer to process, but offer more information and thereby allow the neural network to make better decisions.

[0051] The neural network is not limited to operating on grayscale images. The neural network can be applied to color images in several ways, including but not limited to the following. As a first example, the color image is converted from non-perceptual color space (e.g., RGB) to a perceptual color space such as YCC or LAB, and only the luminance information is modified by the neural network. The color channels are not processed for artifact reduction. The modified luminance and unmodified color channels are converted back to the non-perceptual color space.

[0052] In the next three examples, artifacts are removed from the color channels. In JPEG, the color channels tend to be compressed much more than the luminance channel, mainly because the human visual system is not as sensitive to color shifts. However, blocked colors can be visually disturbing, so modifying the color channels with the neural network to smooth block edges can be beneficial.

[0053] As a second example of processing a color image, the input image is given in non-perceptual color space such as RGB. Neural networks are applied separately to each channel. Outputs of the neural networks provide color components of the output image.

[0054] As a third example of processing a color image, the input image is given in non-perceptual color space. The input image is converted to perceptual color space, each channel is modified by a neural network, and the modified channels are converted back to non-perceptual color space.

[0055] As a fourth example, the input image is given as an RGB representation. A luminance channel is extracted from the input image and modified

by the neural network. A modified color image is produced by adding a delta to each channel of each pixel in the non-perceptual color image. Each delta is computed as the difference between the corresponding luminance value in the modified luminance channel and the corresponding luminance value in the original luminance channel.

[0056] If any of the values of the modified image is outside of the RGB gamut, a gamut-clipping operation may be applied. Each RGB value may be clipped separately to the allowed range, or a more sophisticated gamut-mapping method may be used to preserve perceptual attributes such as hue. One such gamut-mapping method is disclosed in U.S. Serial No. 10/377,911 entitled "System and method of gamut mapping image data" and filed on Feb 28, 2003. In this fifth example, a color image is modified without computing chrominance channels explicitly.

[0057] There is no preferred hardware implementation for the method of training the neural networks according to the present invention, and there is no preferred hardware implementation for the trained neural networks. An exemplary hardware implementation for both the training of neural networks and a trained neural network is illustrated in Figure 6.

[0058] Referring to Figure 6, a computer 610 includes a processor 612 and computer memory 614. The memory 614 stores the details of the trained neural network 616, including information about the weights of the input, hidden and output nodes.

[0059] The neural network 616 may be trained and used in the same computer 610, or it may be trained on one computer and used on one or more other computers. The training computer may store a program 620 and training images 622 for training the neural network 616 in accordance with the method described above. The trained neural network 616 can then be transmitted other computers.

[0060] The neural network 616 can be distributed (e.g., sold commercially) in any number of ways. For example, the neural network can be distributed via a removable medium 618 such as an optical disc (e.g., DVD) or transmitted (e.g., as an installable package over the Internet) from memory of one machine to another. The

neural network 616 could be loaded on a machine (e.g., a printer, a personal computer), which would be distributed.

[0061] The output image(s) can also be distributed in any number of ways. For example, the output images can be distributed via a removable medium 718 such as an optical disc (e.g., DVD) or transmitted (e.g., over the Internet) from memory of one machine to another.

[0062] The trained neural network is not limited to any particular application. An exemplary application is a printer pipeline. If an image to be printed is a strongly compressed JPEG image, then it would be desirable to reduce the artifacts before performing other image processing, such as denoising, sharpening, and contrast enhancement. A controller of the printer could be programmed with a neural network for performing the artifact reduction.

[0063] As mentioned above, the present invention is not limited to JPEG images. The neural network according to the present invention could be used with a variety of transform methods, such as JPEG2000 and PNG. It could even be applied on a per-frame basis to MPEG and other similar algorithms.

[0064] The present invention is not limited to the specific embodiments described above. Instead, the present invention is construed according to the claims the follow.